

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## The use of artificial neural networks to classify primate vocalizations: a pilot study on black lemurs

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/71122> since

*Published version:*

DOI:10.1002/ajp.20786

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

## RESEARCH ARTICLE

# The Use of Artificial Neural Networks to Classify Primate Vocalizations: A Pilot Study on Black Lemurs

LUCA POZZI<sup>1–3\*</sup>, MARCO GAMBA<sup>1</sup>, AND CRISTINA GIACOMA<sup>1</sup>

<sup>1</sup>Dipartimento di Biologia Animale e dell'Uomo, Università di Torino, Torino, Italy

<sup>2</sup>Center for the Study of Human Origins, Department of Anthropology, New York University, New York, New York

<sup>3</sup>New York Consortium in Evolutionary Primatology (NYCEP), New York, New York

The identification of the vocal repertoire of a species represents a crucial prerequisite for a correct interpretation of animal behavior. Artificial Neural Networks (ANNs) have been widely used in behavioral sciences, and today are considered a valuable classification tool for reducing the level of subjectivity and allowing replicable results across different studies. However, to date, no studies have applied this tool to nonhuman primate vocalizations. Here, we apply for the first time ANNs, to discriminate the vocal repertoire in a primate species, *Eulemur macaco macaco*. We designed an automatic procedure to extract both spectral and temporal features from signals, and performed a comparative analysis between a supervised Multilayer Perceptron and two statistical approaches commonly used in primatology (Discriminant Function Analysis and Cluster Analysis), in order to explore pros and cons of these methods in bioacoustic classification. Our results show that ANNs were able to recognize all seven vocal categories previously described (92.5–95.6%) and perform better than either statistical analysis (76.1–88.4%). The results show that ANNs can provide an effective and robust method for automatic classification also in primates, suggesting that neural models can represent a valuable tool to contribute to a better understanding of primate vocal communication. The use of neural networks to identify primate vocalizations and the further development of this approach in studying primate communication are discussed. Am. J. Primatol. 72:337–348, 2010. © 2009 Wiley-Liss, Inc.

**Key words:** automatic classification; *Eulemur macaco macaco*; Multilayer Perceptron; vocal repertoire; cluster analysis; discriminant function analysis

## INTRODUCTION

The structural identification of different vocal units constitutes a crucial step for an effective comparison of vocal repertoires. Owing to the need of identifying discrete categories in comparing diversity at several level of analysis (description of different notes to define the vocal repertoire, function- and context-dependent variability, ecologically related diversity, species, and individual identification, etc.), the development of mathematical computational techniques for discrimination among vocal signals has become one of the main tasks in studying animal vocal communication [Clemins et al., 2005; Gamba & Giacoma, 2007; Murray et al., 1998].

Traditionally, within primates, the study of the vocal repertoire has been performed using the following three different approaches: (1) contextual approach that identifies vocalizations emitted in different behavioral situations [Baldwin & Baldwin, 1976]; (2) behavioral approach that looks at vocalizations and behaviors to infer functional properties of the calls [Pereira et al., 1988]; and (3) acoustic

approach that mainly classifies different vocal types on the basis of their acoustic structure with a limited impact of behavioral information [Gamba & Giacoma, 2005, 2007]. A combination of behavioral and acoustic data is now widely used in primatology [Gosset et al., 2002; Gros-Louis et al., 2008; Range & Fischer, 2004] and several statistical methods are commonly used, such as Multivariate ANalysis Of VAriance [MANOVA; Owren et al., 1997; Range & Fischer, 2004], Discriminant Function Analysis [DFA; Gamba & Giacoma, 2007], Principal Components Analysis [PCA, Gros-Louis et al., 2008; Pistorio et al., 2006; Rendall, 2003], and Cluster Analysis [CA; Hammerschmidt & Fischer, 1998;

\*Correspondence to: Luca Pozzi, Department of Anthropology, New York University, 25 Waverly Place, New York, 10003. E-mail: luca.pozzi@nyu.edu

Received 20 February 2009; revised 24 November 2009; revision accepted 29 November 2009

DOI 10.1002/ajp.20786

Published online 23 December 2009 in Wiley InterScience (www.interscience.wiley.com).

Hammerschmidt & Todt, 1995]. An alternative to the statistical approach, but still not common in primatology, is the comparative classification of nonhuman vocalizations by multiple human listeners [Nicastro & Owren, 2003; Riede et al., 2005].

Unfortunately, current classificatory methods do not always give unambiguous results, in particular when data are nonlinearly distributed (i.e. vector inputs that are not separable into two subsets by a plane) [Demuth & Beale, 1993] and, moreover, present some disadvantages. First, most studies focusing on acoustic classification very often involve a number of a priori assumptions that may affect the objectivity of the classification and, therefore, the results of the analyses can be difficult to compare with other studies. Second, most of the statistical techniques mentioned above are time-consuming and do not allow an automatic or semi-automatic analysis of large numbers of calls [Placer & Slobodchikoff, 2000; Placer et al., 2006].

However, until the end of the 1980s, little progress was made in developing a more reliable or “objective” procedure to categorize vocal repertoires in different species. Over the last decade, the development of cheap and fast computers and the availability of reliable, inexpensive software have encouraged researchers to design computational methods of classification. Nowadays, the attention of researchers is focused on establishing new techniques that may provide an independent call categorization, based on the acoustic structure of the vocalization, in order to minimize the effect of the researcher subjective perception and to provide a tool for fast screening of a large number of vocalizations. Automatic systems of vocal repertoire categorization directed toward the recognition of potential discrete vocal categories have recently been explored in several mammal species [African elephants: Clemins et al., 2005; bats: Melendez et al., 2006; humpback whales: Mercado & Kuh, 1998; false killer whales: Murray et al., 1998]. One benefit is that these systems are not only able to classify a large number of signals, but the automatic procedure can also reduce the human bias in the classification process and, therefore, allows replicable results across different studies [Melendez et al., 2006].

Artificial Neural Networks (ANNs) represent one of those automatic systems that have been widely used in the field of pattern recognition and acoustic classification. They provide an effective, yet partially unexplored, method for solving relevant problems in this field. Neural networks grew out of research in artificial intelligence in the 1960s, and they attempt to mimic the fault-tolerance and capacity to learn of biological neural systems by modeling the low-level structure of the brain [Patterson, 1996]. The great success of ANNs can be attributed to their ability of learning from experience in order to improve their performance,

and to adapt themselves to changes in the environment, to better deal with incomplete information or noisy data [Placer & Slobodchikoff, 2000], and to address nonlinear relation problems [Demuth & Beale, 1993].

Here, we illustrate the first application of ANNs for classifying primate vocalizations in discrete categories. Our study has three main goals: (1) to introduce the framework and functioning of ANNs and their use in primatology; (2) to apply ANN to a sample of vocalizations in order to investigate its potential as a classificatory tool of nonhuman primate vocal repertoire; and (3) to compare the performance of ANNs with statistical methods, which are already commonly used in studying primate communication, such as Discriminant Function Analysis and Cluster Analysis.

To achieve these goals, we examined some of the most common vocalization types uttered by black lemurs (*Eulemur macaco macaco*). Like many other primate species, black lemurs show a vocal repertoire including vocalizations that strongly differ in duration, spectral structure, and phonation mechanism [Gamba & Giacoma, 2005; Gosset et al., 2002; Macedonia & Stanger, 1994]. This species is, thus, a very good test bed to investigate the ability of ANNs in categorizing vocal signals. Although the lemurs of Madagascar have been widely studied by primatologists, only a few articles have investigated the modalities of communication in the genus *Eulemur* and the vocal repertoire has been described for a few species only. Pioneering studies by Klotz [1966] and Petter et al. [1977] described some vocalizations emitted by *E. macaco macaco* but, since then, technical advances in bioacoustics have allowed researchers to analyze the black lemurs' vocal repertoire in a more detailed way. In Table I, we present a brief description of seven vocalization classes that have been recognized by earlier studies based on the assessment of spectrograms types with behavioral analysis and statistical classification [Gamba & Giacoma, 2005; Gosset et al., 2002; Macedonia & Stanger, 1994]. An example of the spectrogram for some of these vocalizations is shown in Figure 1.

## A Brief Introduction to ANNs

ANNs are mathematical/computational models based on biological neural networks. They consist of an interconnected group of simple elements or units (neurons), usually organized in different layers (Fig. 2). A set of numeric values (weights and biases) is associated to each interneuronal connection and is adjusted over time to create a mapping between an input matrix and an output space. The operation of a general neural network can be simply divided into two phases: (1) the *learning* and (2) the *classification* phase. During the *learning* phase (or training), the network learns (adjusting weights and biases) to

**TABLE I. Vocal Categories Recognized by Earlier Studies [Gamba & Giacoma, 2005; Gosset et al., 2002; Macedonia & Stanger, 1994] within Black Lemur Vocal Repertoire**

| Vocal category               | Description   | Behavioral context   | References   |
|------------------------------|---|--|--|
| <i>Grunt</i>                 | Low-pitched signal, with a short duration (< 150 ms) consisting of single-unit frequency spikes         | Usually emitted when the individuals are in visual contact with other members of the group       | Gosset et al. [2002]; Macedonia and Stanger [1994]     |
| <i>Long grunt</i>            | Similar to <i>Grunt</i> but with a much longer duration (> 500 ms)                                      | Usually given, by both sexes, when there is a high level of arousal in the group                 | Gamba [2005]; Macedonia and Stanger [1994]             |
| <i>Grunted hoot</i>          | Low-pitched portion similar to the grunt and a low-pitched tonal portion appended at the end            | Multi-contextual function and can be emitted in the same context of emission of grunts           | Macedonia and Stanger [1994]                           |
| <i>Clear call</i>            | Low-amplitude characterized by a high fundamental frequency and a short duration                        | These tonal calls help to maintain contact between individuals                                   | Gosset et al. [2002]; Macedonia and Stanger [1994]     |
| <i>Alarm call</i>            | Long duration and a high fundamental frequency. It is often preceded by a series of introductory pulses | Contexts of environmental disturbance and agitation, also elicited by potential aerial predators | Macedonia and Stanger [1994]                           |
| <i>Hoot</i>                  | Brief and low-pitched, tonal signals  | General contact function and can be emitted in the same context of emission of grunts            | Macedonia and Stanger [1994]                           |
| <i>Long grunt clear call</i> | Various forms of grunts can be appended with tonal elements   | Usually given by both sexes, most often when there is a high level of arousal in the group       | Gamba and Giacoma [2005]; Macedonia and Stanger [1994] |

recognize specific output targets that are provided by an external operator, by successive iterations (epochs) of a specific set of inputs, called training set. In the *classification* phase, or *generalization*, a new, previously unseen set of inputs is classified in the predefined output categories, according to the classification scheme learnt during the earlier phase (Fig. 2). ANNs are adaptive systems that change their structure based on external or internal information that goes through the network, resembling biological neuron systems not only in their architecture (series of units interconnected to form a web), but also in their functioning; these computational systems are in fact able to “learn” from earlier experience and consequently adapt their ability of classification.

Owing to their ability of modeling complex functions and solving problems of classification and regression, ANNs are today considered as a powerful technique in behavioral sciences [Changeux & Dehaene, 1989; Enquist & Ghirlanda, 2005; Ghirlanda & Enquist, 1998]. In bioacoustics, ANNs have been applied to address several classification tasks, such as species [Chesmore, 2001, 2004; Chesmore & Ohya, 2004; Deregnacourt et al., 2001; Houser et al., 1999; Parsons, 2001; Parsons & Jones, 2000; Tian & Shang, 2006] and individual recognition [Campbell et al., 2002; Deecke et al., 1999; Reby et al., 1997], discrimination of different call types [Mellinger & Clark, 2000; Mercado & Kuh, 1998; Murray et al., 1998; Potter et al., 1994; Schon

et al., 2001], and analysis of the call structure [Dawson et al., 2006; Nickerson et al., 2006]; today, many authors consider this approach to yield better results than conventional techniques [Mercado et al., 2008; Placer & Slobodchikoff, 2000; Reby et al., 1997]. ANNs, in fact, present several advantages over classical statistical methods because they do not require any assumptions about both the input characteristics, such as data linearly distributed and the acoustic features that can be relevant for the classification [Mercado et al., 2008].

Although today neural networks are widely applied in biology, to our knowledge, no experimental studies have ever applied this computational technique in studying nonhuman primate vocalizations. Almost 15 years ago, Zimmermann [1995] suggested neural network modeling as a powerful tool for the analysis and interpretation of primate signals; however, since then, no studies have used neural network to investigate acoustic communication in primates, despite the wide application in other mammal and bird taxa. This article is aimed at fulfilling this lacuna and to suggest neural network as a powerful technique also in studying primate vocal communication.

## METHODS

### Sample: Animals and Recordings

The research described here utilized digitized calls made from tape recordings of calls obtained

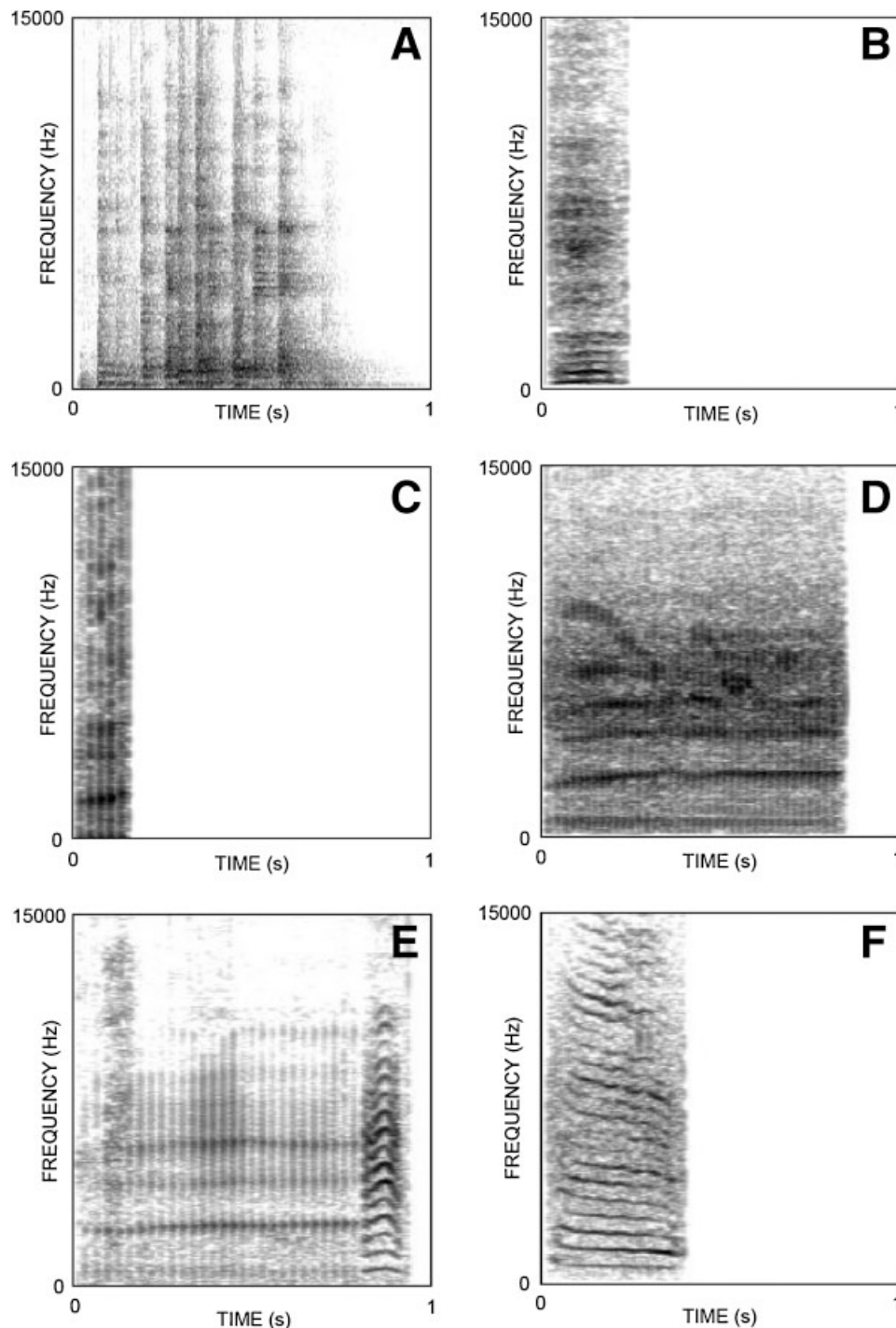


Fig. 1. Spectrograms of some black lemur vocalizations: (A) Alarm Call; (B) Hoot; (C) Grunt; (D) Long Grunt; (E) Long Grunt Clear Call; and (F) Tonal Call.

over a period of 10 years (1995–2005), at several captive and wild populations of black lemurs. All individuals were kept in captivity in zoos in Europe (Parco Natura Viva (PNV), Bussolengo; Verona and Bioparco, Rome), Madagascar (Parc Botanique et Zoologique de Tsimbazaza (PBZT), Antananarivo), the United States (St. Louis Zoo), and in the wild at

the Lokobe Lemur Reserve, Madagascar. In PNV, Bioparco, and St. Louis Zoo, the groups were kept in two indoor–outdoor enclosures, whereas in PBZT, the animals were housed in one cage with an artificial cover. During the day, lemurs had free access both to the indoor and outdoor spaces (when available). The floors in each indoor area were

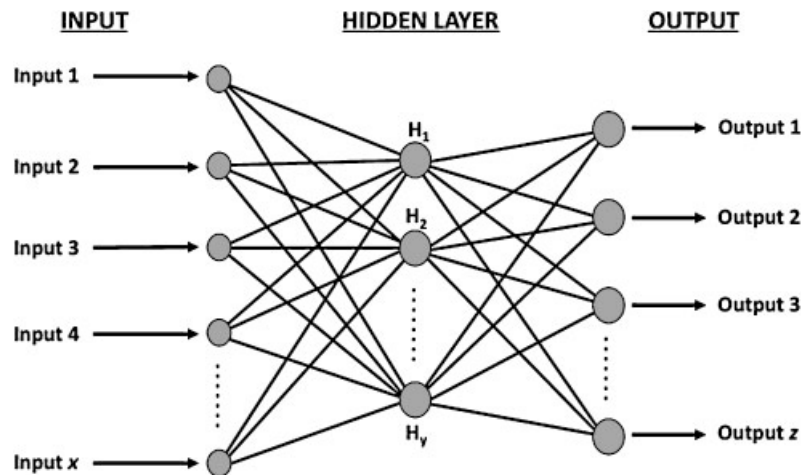


Fig. 2. Structure of a one-hidden layer supervised neural network, similar to the multilayer perceptron used in this study.

covered with sawdust, whereas the outdoor areas were enriched by trees, trunks, rocks, and artificial tools, such as ropes and perches. All enclosures were approved by the World Association of Zoos and Aquarium animal welfare standards. Subjects usually received food twice a day; all animals were fed a diet of fruit, vegetables, and seeds, and had access to water ad libitum. All research, reported in this article, adhered to the American Society of Primatologists' principle for the ethical treatment of nonhuman primates and was approved by the Italian Ministry of Scientific Research (MIUR), the University of Torino, and all zoological parks where the recordings were performed.

Recordings were obtained by using Digital Audio Tape (DAT; e.g. Sony TCD-D100) recorders or solid-state recorders (e.g. Marantz PMD671) and ultra-directional microphones, as Sennheiser ME88 with K3U module. The database includes calls from 26 individual black lemurs. A total of 344 vocalizations were selected from the recordings described above, including 20 different individuals, twelve males and eight females. The final data set included seven different vocal types: 28 *Alarm Calls*, 22 *Grunted Hoots*, 31 *Hoots*, 156 *Grunts*, 23 *Long Grunts*, 44 *Long Grunt Clear Calls*, and 40 *Tonal Calls*. Vocalizations were randomly sampled from the recordings; however, spectrograms for each signal were visually evaluated and poor quality signals were excluded from the final sample. All vocalizations were visually or acoustically identified and then saved as single files onto a computer via sound cards (e.g. Creative Labs SoundBlaster Audigy), at a sampling rate of 44,100 Hz using the software Sony SoundForge version 7.0.

### Acoustic Analyses and Construction of the Input Vectors

We classified each signal in a specific vocal category following some earlier classifications of

black lemur vocal repertoire [Gamba & Giacoma, 2005; Gosset et al., 2002; Macedonia & Stanger, 1994]. This preliminary classification, necessary to test the efficiency of different methods, was based on the spectrogram assessment and acoustic discrimination by different researchers. Both ANNs and statistical methods have been tested using the same sample as described above. In order to fully automatize the sound analysis, we set up an automatic procedure that allowed us to describe each vocalization on the basis of both spectral (fundamental frequency and formants) and temporal (duration) parameters. Sound features were automatically extracted from each sound using the software PRAAT [Boersma & Weenink, 2005] combined with AKUSTYK, a comprehensive vowel analysis software package by Plichta [2005] at the Michigan State University. To detect fundamental frequency, Fast Fourier transforms were generated for all calls (frequency range: 015,000 Hz; maximum: 50 dB/Hz; dynamic range: 30 dB; pre-emphasis: 6.0 dB/Oct; dynamic compression: 0.0). Formants were studied using linear predictive coding (LPC). Two methods were used to evaluate the formant prediction by LPC. First, formant analyses were superimposed over the signal spectrogram. Second, autocorrelation-based LPC spectra were overlaid on independently derived Fast Fourier Transform spectra of the same frames to verify the LPC analysis. Typical settings were 12 kHz for the maximum formant and 22.0 Hz for the dynamic range. The formant pattern was inferred during a step-by-step monitored process, where the operator could interrupt the analysis and modify the analysis parameters [see also Gamba & Giacoma, 2005, 2007, for further details about acoustic features extraction].

For each signal, the software saved a spreadsheet summarizing all spectral and temporal measurements extracted; then, each signal was automatically transformed in an input vector and all vectors were

joined in a matrix ( $13 \times 311$  elements). The neural network analyses were performed using MATLAB 7.0 [Matlab Neural Network Toolbox: Demuth & Beale, 1993], STATISTICA NEURAL NETWORKS 4.0 (StatSoft, Trajan Software Ltd, 1996–2000), and WEKA 3 [Witten & Frank, 2005].

This automatic procedure of acoustic feature extraction allowed us to describe each sound by taking into account not only the acoustic/physical structure of each vocalization, but also to detect how these features change over time. The characterization of acoustic signals in input vectors is a crucial point in neural network analysis, and earlier studies have addressed this problem in different ways. One of the most popular ways to construct input vectors has been the codification of time–frequency representations or spectrograms [Deregnacourt et al., 2001; Walker et al., 1996]. However, spectrograms may not be appropriate to characterize objectively acoustic signals because the time–bandwidth trade-off can dramatically affect the representation of the signal. Slight changes in the window size may in fact result in very different spectrogram representations

of the same signal and, therefore, an arbitrary choice of window sizes may mislead categorization [Murray et al., 1998]. Moreover, to include sound of different duration, spectrograms have to be stretched to create vectors of the same length [Deregnacourt et al., 2001]. This procedure implies a distortion of signals that strongly differ in duration and may not provide an “objective” representation of the call structure. The characterization of sounds, using spectral and temporal parameters, has been proved to provide an accurate description of primate vocalizations [Fitch & Hauser, 2002; Gamba, 2005; Owren & Rendall, 2001]. In order to describe the time-variable features, each sound was, therefore, characterized by spectral measurements (fundamental frequency and formants) at three different points along sound duration (beginning, center, and end of the signal). Each vocalization was thus characterized by a 13-element vector: three elements each corresponding to measurements of fundamental frequency (F0) and first three formants (F1, F2, and F3) and one element corresponding to duration. Figure 3 summarizes how each signal was characterized in an

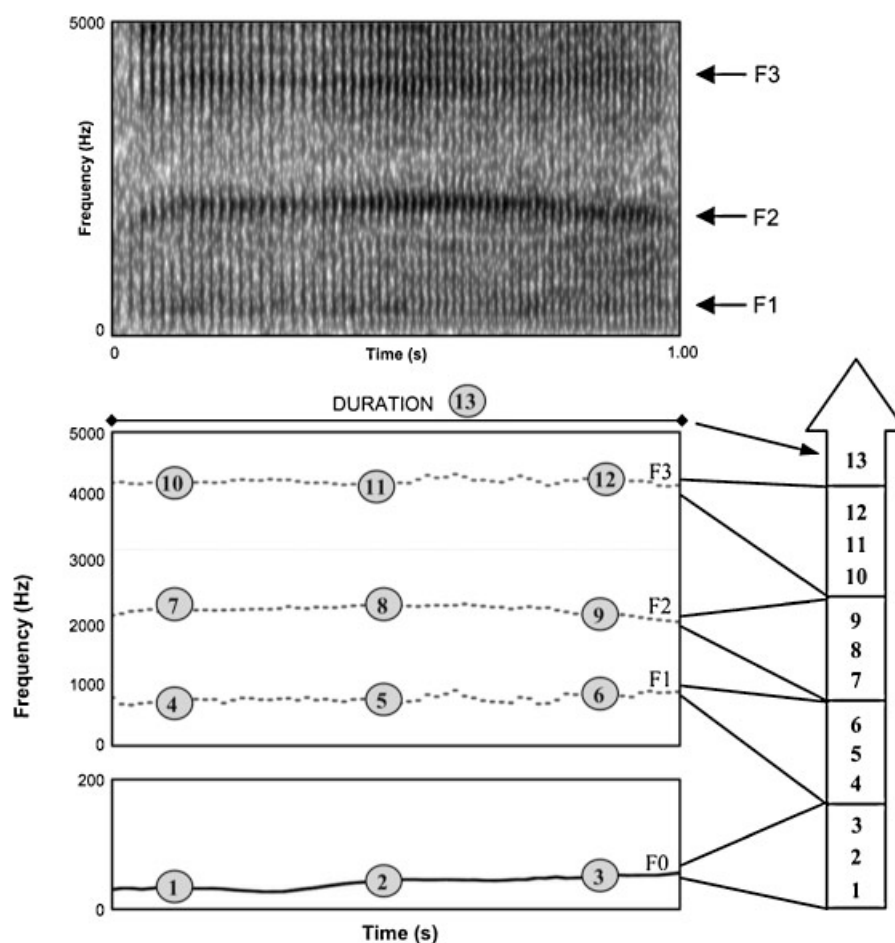


Fig. 3. Structure of the input used in the study. We collected 13 measurements to build each input vector used in this study. 1–3: fundamental frequency (F0); 4–6 first formant (F1); 7–9: second formant (F2); 10–12: third formant (F3); 13: duration (DUR).

input vector. This approach allowed us to describe the array of vocal types uttered by black lemurs, independently from the duration or the acoustic structure of the sound. Additionally, in contrast to some earlier studies, the total duration of signals have been used and all parameters have been sampled using the same time step window, avoiding distortion of sounds.

A full data set (including all 344 signals) and a reduced subset (only 232 signals) were built, in order to test the classification ability of both ANNs and statistical analyses where signals are unevenly distributed across different categories. In our case, the number of *Grunts* exceeds the average of the other six categories; therefore, a subset where the number of these signals was randomly reduced to 44 (same amount of the second most numerous category, *Long Grunt Clear Call*) was employed for each analysis.

## Experimental Protocol

### *Artificial neural network*

The input sample (full data set or subset) was divided into two sets: a training set (67% recordings) and a test set (33%). In each of the analysis we performed, the two sets were randomly subsampled, with respect to the fraction of each vocal type, 10 times. Moreover, we subsampled the whole set into 10 different subsets, submitted to additional runs of the analyses. To reduce the risk of overlearning (when the network overfits the training data and cannot generalize previously unseen data during the test phase), we applied a cross-verification approach. To determine the structure of the network and then to test the reliability of the results obtained, we followed the same experimental protocol described by Reby et al. [1997]. We first performed a series of analyses in order to assess the network behavior, varying the number of neurons and iterations during the training and the test phase. These preliminary analyses allowed us to choose the best network architecture, i.e. the network with the best performance.

### *Statistical analyses*

Several statistical analyses have been commonly used in studying nonhuman primate vocal repertoires, including Stepwise Discriminant Function Analysis (sDFA) and Cluster Analysis (CA) [Fischer et al., 2001; Gouzoules & Gouzoules, 1989; Hammerschmidt & Fischer, 1998; Hammerschmidt & Todt, 1995]. In order to test the ability of ANNs in classifying discrete acoustic categories, we applied both sDFA and CA on the same data set described above (13 variables: three for F0, three for F1, F2, F3, respectively, and one for duration). Specifically, sDFA has been applied to identify linear combinations of predictor variables that maximize the

differences among vocal types [Lehner, 1998]. sDFA is a valuable tool to define which variable(s) discriminates best between groups (e.g. vocal types, species, or individuals). Among the measurements taken from the various representations of a vocal signal, it is possible to show which ones are useful to discriminate between groups, and then to combine them into a number of classification scores ( $n-1$ ). The sDFA was run using a stepwise procedure, and  $F$ -value thresholds for acceptance or rejection of independent variables were set at  $F = 2.71$  and  $3.84$ , in all analyses. First, sDFA was run on both the full data set and the subset described above. To assess to what extent differences among individuals might have biased the statistical inferences, we repeated sDFA on 20 randomly selected subsamples, each of them composed of 70% of the complete vocalization set [Mundry & Sommer, 2007]. The mean percentage of correct classification was considered as an estimate of the robustness of the analysis.

We performed a K-means cluster analysis to confirm the presence of seven call types corresponding to our own subjective classification. K-means clustering seeks to partition the data set into a predefined number of groupings to minimize variability within clusters and maximize variability between clusters. Cluster analysis has been run for 100 iterations and minimum standard deviation of  $10^{-6}$ .

## RESULTS

### **Artificial Neural Networks**

During the preliminary analyses, the behavior of the network according to the number of epochs and interactions was studied using a trial-and-error approach; we selected a 15 units network (one hidden layer network; Fig. 3), trained for 500 iterations (learning rate = 0.2; momentum = 0.2), which yield the best performance with 93.7% of correct attributions for the full data set and 94.0% for the subset test phase.

Because we initialized our ANN with random connection weights before the training, the results of different training can vary based on both the training sample and the initial assignation of connection weights. This particular aspect results in a possible differential ability of separately trained network to classify input signals. To test the reliability and the consistency of the network classification independently from the weight initialization, we tested ten experimental runs, each performed with an independently trained network, and ten different training and test sets by randomly sampling the vocalizations from the whole data set. We then analyzed the mean performance for each target output (vocal category) in the ten runs. The results for each category in the ten runs are reported in Table II. Different training networks showed



**TABLE II. Correct Classifications for Each of the Seven Target Categories in a Supervised Neural Network with 15 Hidden Units**

| Run      | Alarm | Grunted hoot | Grunt | Hoot | Long grunt | Long grunt clear call | Tonal | Total performance |
|----------|-------|--------------|-------|------|------------|-----------------------|-------|-------------------|
| 1        | 89.3  | 81.8         | 90.2  | 96.8 | 100.0      | 95.7                  | 100.0 | 94.1              |
| 2        | 92.9  | 86.4         | 71.4  | 96.8 | 100.0      | 91.3                  | 97.5  | 92.8              |
| 3        | 89.3  | 81.8         | 84.6  | 93.5 | 100.0      | 95.7                  | 100.0 | 93.5              |
| 4        | 89.3  | 81.8         | 86.7  | 90.3 | 100.0      | 91.3                  | 100.0 | 93.1              |
| 5        | 85.7  | 81.8         | 91.4  | 90.3 | 100.0      | 95.7                  | 100.0 | 93.3              |
| 6        | 89.3  | 81.8         | 81.8  | 90.3 | 100.0      | 100.0                 | 100.0 | 93.3              |
| 7        | 82.1  | 81.8         | 88.5  | 90.3 | 100.0      | 95.7                  | 100.0 | 92.5              |
| 8        | 85.7  | 86.4         | 88.5  | 96.8 | 100.0      | 95.7                  | 100.0 | 94.4              |
| 9        | 96.4  | 86.4         | 87.5  | 90.3 | 100.0      | 95.7                  | 100.0 | 94.8              |
| 10       | 92.9  | 86.4         | 93.8  | 93.5 | 100.0      | 95.7                  | 100.0 | 95.6              |
| Average  | 89.3  | 83.6         | 86.4  | 92.9 | 100.0      | 95.2                  | 99.8  | 93.7              |
| St. Dev. | 4.1   | 2.3          | 6.3   | 3.0  | 0.0        | 2.5                   | 0.8   | 1.0               |

Each value is averaged across the ten different datasets we used to test the reliability of the network classification.

variations in performance values. The overall performance of the network varied from 92.5 to 95.6% for the entire data set. In total, all seven categories were recognized by the network with percentage of correct classification ranging from 83.6% for the *Grunted Hoot* to 100% for the *Long Grunt* (Table II). In all ten runs performed, seven vocal types were correctly classified with percentage over 80%. Only in run two, *Grunts* were recognized with lower performance (71.4%). Most cases signals that were classified differently from the a priori categories were misclassified among the categories *Grunt* and *Grunted Hoot*. The overall average performance of prediction of the network, i.e. the number of signals that were assigned to the correct category, was 93.7% (Table II).

### Statistical Analyses: sDFA and Cluster Analysis

The sDFA, used to discriminate among vocal types, showed Factor 1 primarily based on F02 (loading: -4.3719) and F03 (loading: 3.48) while Factor 2 was primarily based on F02 (loading: 1.894) and Dur (loading: -0.806). The sDFA model correctly categorized 85.7% *Alarm Calls*, 90.9% *Grunted-Hoots*, 79.5% *Hoots*, 74.2% *Long Grunts*, 97.7% *Long Grunt Clear Calls*, 95.7% *Grunts*, and 100% *Tonal Calls*. The Discriminant Model was highly significant and correctly classified 86% (83.7% cross-validated jackknife classification) of cases according to the vocal type. There were six Discriminant dimensions, five of which were statistically significant. The canonical correlations for the first two dimensions were 0.97 and 0.96, respectively (Wilks' Lambda = 0.083;  $F_{12,672} = 138.756$ ,  $P < 0.001$ ) in the full data set. The sDFA applied to the 20 subsets showed average correct classification of  $87.3 \pm 2.2\%$  of the cases.

Cluster analysis correctly classified only 76.1% of the signals in the subset, but showed a better

performance than sDFA in the full data set (88.4%). The performance for each of the vocal types used in the full data set is: 35.7% *Alarm Calls*, 0% *Grunted-Hoots*, 100% *Hoots*, 100% *Long Grunts*, 100% *Long Grunt Clear Calls*, 100% *Grunts*, and 100% *Tonal Calls*.

### Comparative Evaluation of the Classification Rates

Within the same data set, we tested whether the classification rate between Cluster analyses, sDFAs and ANNs, were significant or not. Both sDFAs and ANNs correct classification rates resulted in being higher than those showed by Cluster analyses over the same data set ( $N = 11$ ;  $t = 5.755$ ,  $P < 0.001$  and  $N = 11$ ;  $t = 10.001$ ,  $P < 0.001$ , respectively). In the comparison between sDFA and ANN, the result showed that the percentages of correct classification were significantly higher in the networks ( $N = 11$ ;  $t = 10.363$ ,  $P < 0.001$ ).

### DISCUSSION

The study of animal communication is a challenging task for biologists and often requires the development of classification tools in order to categorize vocal signals at different levels of analysis—individuals, populations, and species [Giacoma & Castellano, 1999]. The classification of different vocalizations is not only a preliminary, necessary task in bioacoustic studies, but it is also particularly useful in comparative approach. The comparison of vocal systems across different species requires the identification of homologous signals that can be identified, either studying the behavioral context in which they are emitted or analyzing the acoustic structure of similar signals, or better, combining both approaches. The characterization of animal vocalizations is not trivial and the great variety of vocal types within animal repertoire requires attention by researchers. Moreover, animal vocalizations

are usually given in less standardized conditions and research data sets rarely match speech databases in terms of completeness of information about instances, attributes, and individuals. Therefore, over the last few years, much effort has been directed toward the development of “objective,” automatic, and more effective systems for vocal classification [Clemins et al., 2005; Melendez et al., 2006; Murray et al., 1998].

Although artificial neural networks are now widely used in the study of vocal communication across a high number of taxa [Chesmore, 2001; Dawson et al., 2006; Deregnaucourt et al., 2001; Murray et al., 1998; Parsons & Jones, 2000; Reby et al., 1997], this study represents the first use of this computational tool for the identification of the vocal repertoire in a nonhuman primate species. This is probably owing to the fact that most primatologists are not familiar with this computational method, and because ANNs were considered for a long time as a sort of “black box” to evaluate classification functions. This article was designed to further explore the ability of this approach in identifying discrete categories in primate vocalizations and to compare ANNs with statistical techniques, commonly used in primatology.

Our results demonstrated that ANNs using spectral (fundamental frequency and formants) and temporal (duration) inputs, are a very effective way of categorizing lemur vocalizations. Earlier studies on primate vocalizations have demonstrated that features related to the morphology of the vocal tract (formants) are crucial in discriminating call structure [Fitch, 1997; Owren & Rendall, 2001; Rendall et al., 1998, 2004]. The use of features correlated to anatomical or morphological characteristics as well as physiological or behavioral factors, are therefore to be preferred to others, whose causes of variability are completely unknown. In this study, we combined spectral (fundamental frequency and formants) and temporal features in order to describe as best as possible the acoustic structure of black lemur calls. Earlier studies on lemur vocalizations have in fact shown the importance of duration as a critical feature in distinguishing different calls [Gamba & Giacoma, 2005, 2007]. For example, all earlier studies on the black lemur vocal repertoire identified the *Grunt* and the *Long Grunt* as two different call types emitted in different behavioral contexts [Gosset et al., 2002; Macedonia & Stanger, 1994]. However, the spectral profile of these two calls is identical and they differ only for their duration (<150 ms for grunts and >500 ms for long grunts [Gamba, 2005]). Including also temporal parameter along spectral measurements plays therefore a crucial role in categorizing lemur vocal repertoire. The characterization of vocalizations in this study provides a simple, fast, and reliable description of the call structure, and allows the application of

automatic systems of classification, such as neural networks.

The neural network, applied in this study, showed response properties highly consistent with all seven of the categories developed by human observers [Gamba & Giacoma, 2005; Gosset et al., 2002; Macedonia & Stanger, 1994]. ANNs were in fact able to classify black lemurs' vocalization types with a level of accuracy higher than any other techniques used so far (general prediction of approximately 94%). These results are even more relevant if compared with the performance of the two statistical analyses that are commonly used in the literature to identify different vocal types within nonhuman primate vocal repertoires. The ANNs, in fact, outperformed their equivalent DFA (subsets: 87.3%; full data set: 86.0%) and CA (subsets: 76.1%; full data set: 88.4%). Cluster analysis seems to be more affected by the composition of the sample and its performance differs significantly between the full data set and the subset, with 88.4% and 76.1%, respectively. In contrast, ANNs are not particularly affected by the different composition of the input, and even when signals are unevenly distributed across categories (e.g. *Grunts*, in this study), the overall performance of the network is pretty consistent (subset: 93.7% vs. full data set: 94%).

The comparison between ANNs and statistical analyses performed in this study allows us also to draw some interesting conclusions on the advantages and constraints of the neural network application in primate call classification. As described above, supervised networks need the external operator to define target categories and to train the network to recognize them. Because these ANNs are able to detect autonomously a rule of classification during the learning phase and to reduce the noise in the signals, this approach is extremely powerful to classify not only vocalizations in earlier defined categories, but also in case of low quality recordings [Placer & Slobodchikoff, 2000]. The main advantage of ANNs is to be trained on a well-defined set of signals and then, once the weight factors are saved, easily be reused to classify previously unseen additional records. These ANNs are, thus, better suited to classification tasks where some knowledge of patterns of use is already known. This characteristic makes ANNs a suitable tool for automatic classification of calls: once the network has learnt the classification scheme during the training, it is possible to completely automatize the procedure of call classification, reducing the time of analysis for large data set.

Although we encourage the use of ANNs in studying vocal communication in nonhuman primates, we also acknowledge some of the limitations of this approach. First, the performance of ANNs may be influenced by the training stage. Both the composition of the training set and the protocol

applied may in fact reduce the ability of generalization of the network, affecting the overall performance during the test phase [Ozesmi et al., 2006]. For instance, if the signals used in the training set are from only one or a few animals and do not well represent the interindividual variability, ANNs may “learn” to recognize some acoustic features that are not representative of the species itself (or more specifically, are individual- or sex-specific). Moreover, if the training phase is not run properly, the risk of overgeneralization may be common. In this situation, the network can overfit the training data (*overlearning*), resulting in a network able to perfectly recognize the training sets, but not to generalize previously unseen data during the test phase. This problem can be easily overcome, running a network with different training sets and different architecture in order to assess the behavior of the network itself and to avoid overlearning on a specific data set. Moreover, the use of a verification set (as employed in this study) can reduce the risk of overlearning [Demuth & Beale, 1993].

A second major problem of supervised ANNs is that they require the definition of a priori categories during the learning phase and, therefore, not helpful in establishing new categories. Although ANNs, such as the multilayer perceptron used in this study, can classify signals only in a predefined number of categories, there are other network models that may be applied to recognize new categories within a vocal repertoire. For example, competitive self-organizing maps [Kohonen, 1988] can be used to detect regularities and to classify inputs into discrete categories without defining a priori number and types of target outputs [Murray et al., 1998; Placer et al., 2006]. Although not shown in this study, preliminary analyses on the data set described above using self-organizing networks provided results comparable to the categories used in this study [Pozzi et al., 2006, in press].

To conclude, although our results are promising and demonstrate that ANNs may provide a new powerful tool for studying nonhuman primate vocal communication and more specifically for classifying calls within a vocal repertoire, they should not be generalized. The performance we achieved is particular to the species studied, the methods with which their calls were recorded and analyzed, and the variables we chose to measure. However, our study suggests that ANNs may not only be more powerful in performing classification task than equivalent statistical analyses, but also be very helpful to assign calls with a high background noise. This is especially relevant because most primates live in forest habitats where background noise is particularly high and good quality recordings are often hard to obtain (especially for poor-studied species). Further development of this approach will include the application of ANNs to the classification of vocalizations within

other species of nonhuman primates and also their use to address other problems in primate bioacoustics, such as individual, sex, or species identification [Pozzi et al., 2008, in press].

## ACKNOWLEDGMENTS

We are grateful to Sergio Castellano for his help in MATLAB analyses and for advice and support throughout the preparation of this article. We thank Stuart Parsons for his technical comments about the use of neural networks in bioacoustics, Caroline Harcourt, Paolo Cocco, and two anonymous reviewers for their critical comments and suggestions on early drafts of this article. We also thank Cesare Avesani Zaborra and Parco Natura Viva—Garda Zoological Park for their continued support.

## REFERENCES

- Baldwin JD, Baldwin JI. 1976. Vocalizations of Howler monkeys (*Alouatta palliata*) in Southwestern Panama. *Folia Primatologica* 26:81–108.
- Boersma P, Weenink D. 2005. PRAAT: doing phonetics by computer: <http://www.praat.org>
- Campbell GS, Gisiner RC, Helweg DA, Milette LL. 2002. Acoustic identification of female Steller sea lions (*Eumetopias jubatus*). *Journal of the Acoustical Society of America* 111:2920–2928.
- Changeux JP, Dehaene S. 1989. Neuronal models of cognitive functions. *Cognition* 33:63–109.
- Chesmore D. 2004. Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciências* 76: 435–440.
- Chesmore ED. 2001. Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics* 62:1359–1374.
- Chesmore ED, Ohya E. 2004. Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. *Bulletin of Entomological Research* 94:319–330.
- Clemins PJ, Johnson MT, Leong KM, Savage A. 2005. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *Journal of the Acoustical Society of America* 117:956–963.
- Dawson MR, Charrier I, Sturdy CB. 2006. Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *Journal of the Acoustical Society of America* 119:3161–3172.
- Deecke VB, Ford JK, Spong P. 1999. Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America* 105:2499–2507.
- Demuth H, Beale M. 1993. Neural network toolbox. Natick, MA: The MathWorks, Inc.
- Deregnacourt S, Guyomarc’h JC, Richard V. 2001. Classification of hybrid crows in quail using artificial neural networks. *Behavioural Processes* 56:103–112.
- Enquist M, Ghirlanda S. 2005. Neural networks and animal behavior. Princeton, NJ: Princeton University Press. 253p.
- Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM. 2001. Acoustic features of female chacma baboon barks. *Ethology* 107:33–54.
- Fitch WT. 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America* 102:1213–1222.
- Fitch WT, Hauser MD. 2002. Unpacking “honesty”: vertebrate vocal production and the evolution of acoustic signals.

- In: Simmons A, Fay RR, Popper AN, editors. Springer handbook of auditory research. New York: Springer. p 65–137.
- Gamba M. 2005. Evoluzione della comunicazione vocale nei lemuri del Madagascar. PhD thesis Torino: Università degli Studi di Torino. 328p.
- Gamba M, Giacoma C. 2005. Key issues in the study of primate acoustic signals. *Journal of Anthropological Sciences* 83:61–87.
- Gamba M, Giacoma C. 2007. Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution* 19:323–343.
- Ghirlanda S, Enquist M. 1998. Artificial neural networks as models of stimulus control. *Animal Behaviour* 56:1383–1389.
- Giacoma C, Castellano S. 1999. Advertisement call variation and speciation in the *Bufo viridis* complex. In: Ryan MJ, editor. Recent advances in anuran communication. Washington, London: Smithsonian Institution Press. p 205–219.
- Gosset D, Fornasieri I, Roeder JJ. 2002. Acoustic structure and contexts of emission of vocal signals by black lemurs. *Evolution of Communication* 4:225–251.
- Gouzoules H, Gouzoules S. 1989. Design features and developmental modification of pigtail macaque, *Macaca nemestrina*, agonistic screams. *Animal Behaviour* 37:383–401.
- Gros-Louis JJ, Perry SE, Fichtel C, Wikberg E, Gilkenson H, Wofsy S, Fuentes A. 2008. Vocal repertoire of *Cebus capucinus*: acoustic structure, context, and usage. *International Journal of Primatology* 29:641–670.
- Hammerschmidt K, Fischer J. 1998. The vocal repertoire of Barbary macaques: a quantitative analysis of a graded signal system. *Ethology* 104:203–216.
- Hammerschmidt K, Todt D. 1995. Individual-differences in vocalizations of young Barbary macaques (*Macaca sylvanus*)—a multi-parametric analysis to identify critical cues in acoustic signaling. *Behaviour* 132:381–399.
- Houser DS, Helweg DA, Moore PW. 1999. Classification of dolphin echolocation clicks by energy and frequency distributions. *Journal of the Acoustical Society of America* 106:1579–1585.
- Klotz M. 1966. Lautgebung bei primaten. Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin. Mathematisch-Naturwissenschaftliche Reihe 15:23–56.
- Kohonen T. 1988. The “neural” phonetic typewriter. *Computer* 21:11–22.
- Lehner PN. 1998. Handbook of ethological methods. Cambridge: Cambridge University Press.
- Macedonia JM, Stanger KF. 1994. Phylogeny of the Lemnidae revisited—evidence from communication signals. *Folia Primatologica* 63:1–43.
- Melendez KV, Jones DL, Feng AS. 2006. Classification of communication signals of the little brown bat. *Journal of the Acoustical Society of America* 120:1095–1102.
- Mellinger DK, Clark CW. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America* 107:3518–3529.
- Mercado III E, Kuh A. 1998. Classification of humpback whale vocalizations using a self-organizing neural network. *Proceedings of the IJCNN* 1584–1589.
- Mercado E, Green SR, Schneider JN. 2008. Understanding auditory distance estimation by humpback whales: a computational approach. *Behavioural Processes* 77:231–242.
- Mundry R, Sommer C. 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. *Animal Behaviour* 74:965–976.
- Murray SO, Mercado E, Roitblat HL. 1998. The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America* 104:3626–3633.
- Nicastro N, Owren MJ. 2003. Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology* 117:44–52.
- Nickerson CM, Bloomfield LL, Dawson MRW, Sturdy CB. 2006. Artificial neural network discrimination of black-capped chickadee (*Parus atricapillus*) call notes. *Journal of the Acoustical Society of America* 120:1111–1117.
- Owren MJ, Rendall D. 2001. Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology* 10:58–71.
- Owren MJ, Seyfarth RM, Cheney DL. 1997. The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions. *Journal of the Acoustical Society of America* 101:2951–2963.
- Ozesmi SL, Tan CO, Ozesmi U. 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications 195:83–93.
- Parsons S. 2001. Identification of New Zealand bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) in flight from analysis of echolocation calls by artificial neural networks. *Journal of Zoology* 253:447–456.
- Parsons S, Jones G. 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *Journal of Experimental Biology* 203:2641–2656.
- Patterson DW. 1996. Artificial neural networks. Theory and applications. Singapore: Prentice Hall.
- Pereira ME, Seeligson ML, Macedonia JM. 1988. The behavioral repertoire of the black-and-white ruffed lemur, *Varecia variegata variegata* (Primates, Lemnidae). *Folia Primatologica* 51:1–32.
- Petter JJ, Albignac R, Rumpler Y. 1977. Mammifères lemuriens (Primates prosimiens). Paris: ORSTOM and CNRS.
- Pistorio AL, Vintch B, Wang X. 2006. Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America* 120:1655–1670.
- Placer J, Slobodchikoff CN. 2000. A fuzzy-neural system for identification of species-specific alarm calls of Gunnison’s prairie dogs. *Behavioural Processes* 52:1–9.
- Placer J, Slobodchikoff CN, Burns J, Middleton R. 2006. Using self-organizing maps to recognize acoustic units associated with information content in animal vocalizations. *Journal of the Acoustical Society of America* 119:3140–3146.
- Plichta B. 2005. AKUSTYK: <http://www.msu.edu/~plichta>
- Potter JR, Mellinger DK, Clark CW. 1994. Marine mammal call discrimination using artificial neural networks. *Journal of the Acoustical Society of America* 96:1255–1262.
- Pozzi L, Gamba M, DelPero M, Giacoma C. 2006. Eulemur phylogeny: evidence from molecular and acoustic data. *International Journal of Primatology* 27:434.
- Pozzi L, Gamba M, Giacoma C. 2008. Automated bioacoustic identification of lemur species using artificial neural networks. *Folia Primatologica* 79:123–159.
- Pozzi L, Gamba M, Giacoma C. The use of Artificial Neural Networks in studying lemur vocal communication. In: Masters J, Gamba M, Génin F, editors. Leaping ahead: advances in prosimian biology (Developments in Primatology by Springer Science), in press.
- Range F, Fischer J. 2004. Vocal repertoire of Sooty Mangabeys (*Cercocebus torquatus atys*) in the Tai National Park. *Ethology* 110:301–321.
- Reby D, Lek S, Dimopoulos I, Joachim J, Lauga J, Aulagnier S. 1997. Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes* 40:35–43.
- Rendall D. 2003. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America* 113:3390–3402.
- Rendall D, Owren MJ, Rodman PS. 1998. The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca*

- mulatta*) vocalizations. *Journal of the Acoustical Society of America* 103:602–614.
- Rendall D, Owren MJ, Weerts E, Hienz RD. 2004. Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *Journal of the Acoustical Society of America* 115:411–421.
- Riede T, Mitchell BR, Tokuda I, Owren MJ. 2005. Characterizing noise in nonhuman vocalizations: acoustic analysis and human perception of barks by coyotes and dogs. *Journal of the Acoustical Society of America* 118:514–522.
- Schon PC, Puppe B, Manteuffel G. 2001. Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). *Journal of the Acoustical Society of America* 110:1425–1431.
- Tian H, Shang Z. 2006. Artificial neural network as a classification method of mice by their calls. *Ultrasonics* 44:275–278.
- Walker A, Fisher RB, Mitsakakis N. 1996. Singing maps: classification of whale-song units using a self-organizing feature mapping algorithm. DAI Research Paper No. 833.
- Witten IH, Frank E. 2005. *Data mining: practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann.
- Zimmermann A. 1995. Artificial neural networks for analysis and recognition of primate vocal communication. In: Zimmermann E, Newman JO, Jurgens UWE, editors. *Current topics in primate vocal communication*. New York: Plenum Press. p 29–46.